International Journal of Approximate Reasoning 50 (2009) 315–323

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

Bayesian robustness for decision making problems: Applications in medical contexts

J. Martín [a,*], C.J. Pérez [a], P. Müller [b]

[a] Departamento de Matemáticas, Universidad de Extremadura, Badajoz, Spain
[b] Department of Biostatistics, The University of Texas, M.D. Anderson Cancer Center, Houston, USA

ARTICLE INFO

Article history:
Available online 6 April 2008

Keywords:
Bayesian robustness
Decision making
Markov chain Monte Carlo methods

ABSTRACT

Practical implementation of Bayesian decision making is hindered by the fact that optimal decisions may be sensitive to the model inputs: the prior, the likelihood and/or the underlying utility function. Given the structure of a problem, the analyst has to decide which sensitivity measures are relevant and compute them efficiently. We address the issue of robustness of the optimal action in a decision making problem with respect to the prior model and the utility function. We discuss some general principles and apply novel computational strategies in the context of two relatively complex medical decision making problems.

1. Introduction

Bayesian decision theory describes a decision making problem through a set of feasible alternatives, $a \in \mathscr{A}$; a set of states or parameters, $\theta \in \Theta$; a prior distribution, $\pi(\theta)$; a likelihood, $l(x|\theta)$; and a utility function, $u(a,\theta,x)$. Usually, the utility function does not depend on the data x. In such cases, we write $u(a,\theta)$.

The optimal decision a^* is the alternative that maximizes the posterior expected utility:

$$a^* = \arg \max_{a\in\mathscr{A}} U(a), \tag{1}$$

$$U(a) = \int_\Theta u(a,\theta,x)\pi(\theta|x)\mathrm{d}\theta = \frac{\int_\Theta u(a,\theta,x)l(x|\theta)\pi(\theta)\,\mathrm{d}\theta}{\int_\Theta l(x|\theta)\pi(\theta)\mathrm{d}\theta}. \tag{2}$$

In some cases, the interest focuses on finding an optimal decision before observing data, changing (2) to:

$$U(a) = \int_\Theta u(a,\theta)\pi(\theta)\mathrm{d}\theta.$$

Among many reviews of Bayesian decision theory see, for example, [1–3].

Practical implementation of (1) is hindered by the fact that $U(a)$ and, hence, the optimal action a^* could be sensitive to the chosen prior $\pi(\cdot)$, likelihood $l(\cdot|\cdot)$ and/or utility function $u(\cdot)$. A skeptical decision maker will require, in addition to the optimal solution, some description of its robustness with respect to reasonable changes and imprecisions in the specification of inputs. Moreover, the resolution of (1) must frequently be performed by simulation-based methods, mainly by Markov chain Monte Carlo (MCMC) ones. In these cases, robust analyses may become very involved computationally.

* Corresponding author. Fax: +34 924272911.
E-mail addresses: jrmartin@unex.es (J. Martín), carper@unex.es (C.J. Pérez), pm@odin.mdacc.tmc.edu (P. Müller).

0888-613X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ijar.2008.03.017

Reviews of this area are provided in [4–6]. Most authors, see e.g. [4,7], focus on prior robustness. Sensitivity with respect to the prior and the likelihood is studied in [8,9]. Sensitivity with respect to the loss function is considered in [10,11]. Joint sensitivity with respect to the utility function and the prior is investigated in [12]. Discussions concerning applications to medical models, as the ones we shall pursue here, may be found in [7,13,14].

In this paper, we address the robustness of the optimal action in a decision making problem with respect to the prior model or the utility function. We discuss several general principles and apply new computational strategies in the context of two relatively complex medical decision making problems. The outline of the paper is as follows. In Section 2, we discuss some general principles and new computational strategies to address a robust analysis with respect to the prior distribution or the utility function. Some theoretical results are provided. In Section 3, the proposed robustness analysis is applied to two medical decision making problems, that are representative of a wide range of applications in that area. Section 4 provides conclusions. An Appendix includes proofs of the theoretical results. We emphasize that the developed methodology is potentially applicable to many decision making problems and not just to those described through this paper.

2. Bayesian robustness approach

In this section, $U(a)$ is represented as $U(a;u,\pi)$ because of the dependence relationship. Two key sensitivity aspects (see, for example, [15]) in a decision making problem are:

1. *Expected utility sensitivity*. Changes in the expected utility of a^* as a function of (u,π). It should be, therefore, undertaken a study of the operator:
$$(u,\pi) \to U(a^*;u,\pi). \tag{3}$$
2. *Decision sensitivity*. Changes in the optimal decision a^* if we modify (u,π). Let b be an alternative solution to a^*. If for some (u,π) within a reasonable range of the original assessment, $U(a^*;u,\pi) - U(b;u,\pi) < 0$, then, action b is preferred to the incumbent optimal action a^*. The operators of interest include:
$$(u,\pi) \to U(a^*;u,\pi) - U(b;u,\pi), \tag{4}$$
which expresses differences in expected utility between a^* and b when (u,π) changes;
$$(u,\pi) \to \arg\max_a U(a;u,\pi), \tag{5}$$
which expresses how the optimal alternative changes as (u,π) changes, and
$$(u,\pi) \to U(a^*;u,\pi) - \max_{b\in\mathscr{A}} U(b;u,\pi), \tag{6}$$
which expresses maximum differences in expected utility between a^* and the set of alternatives $b \in \mathscr{A}$, as (u,π) changes. When the value of this operator is 0, a^* remains optimal for any (u,π).

In this paper, we focus on sensitivity with respect to the prior distribution, by considering the operator given in (3), and the operators given by (4) and (5) when the utility function changes.

2.1. Sensitivity with respect to the prior distribution

Many papers on Bayesian robustness have studied this question only focusing on changes in the prior distribution π, allowing π to range in various classes Γ of probability distributions. See, for example, [4,5,16,17] and references therein.

In order to quantify local sensitivity with respect to prior changes, [18] proposed to consider the derivative $\dot{U}_\pi$ of U with respect to π (given a and u), defined as the linear operator such that:

$$U(a;u,\pi+h) = U(a;u,\pi) + \dot{U}_\pi(h) + o(\|h\|) \quad \text{as } \|h\| \to 0,$$

where h is a 0 mass signed measure, and $\|h\|$ is the bounded variation norm:

$$\|h\| = \sup_{B\in\mathscr{B}} |h(B)|.$$

When the decision problem does not include data to update the prior distribution, then $\dot{U}_\pi(\cdot) = U(\cdot)$ because $U(\cdot)$ is linear in π. We may think of $\dot{U}_\pi(h)$ as a rate of change of the expected utility if we change the prior probability measure from π to $\pi + h$. The use of such derivatives in robust Bayesian analysis is reported, among others, in [19–22].

Here, we consider changes in expected utility when varying the prior π. Hence, in the following discussion, the decision a^* and the utility u are fixed, and we denote $U(\pi) = U(a^*;u,\pi)$. The practical impact of the computed prior sensitivity measure would be the following: large values suggest the decision maker that any conclusion should not be applicable to values (u,π) far apart from the initial ones. A more careful prior elicitation might be in order. On the other hand, low values imply that the conclusions can be considered reasonably robust against changes in the prior probability.

Assume that the information about the probability distribution allows us to constrain it to a convex class Γ. Starting with a particular prior probability model $\pi \in \Gamma$, we study the impact of changes in π on the expected utility of a^*. We use neighborhoods $\Gamma_\varepsilon \subseteq \Gamma$ of π defined via ε-contaminations, i.e.:

$$\Gamma_\varepsilon = \{\pi' = \pi + \eta(q - \pi),\ q \in \Gamma,\ 0 \leqslant \eta \leqslant \varepsilon\}.$$

These neighborhoods are commonly used in sensitivity analysis, see, e.g., [1,4,7,11].

For a given π and a given neighborhood Γ_ε of alternative prior models π' around π, the following sensitivity measure is defined in [21]:

$$m_\pi^\varepsilon = \sup_{\pi' \in \Gamma_\varepsilon} |\dot{U}_\pi(\pi' - \pi)|.$$

Note that this supremum is an upper bound on the changes in expected utility when varying the prior model infinitesimally in this class as

$$|U(\pi') - U(\pi)| \leqslant \sup_{\rho \in \Gamma_\varepsilon} |\dot{U}_\pi(\rho - \pi)| + \circ(\|\pi' - \pi\|),$$

where $\dot{U}_\pi$ is the derivative of U with respect to π evaluated at $\delta = \pi\prime - \pi$. Due to the fact that $\|\pi' - \pi\| \leqslant \varepsilon\ \forall \pi' \in \Gamma_\varepsilon$, we define an infinitesimal sensitivity measure for the prior distribution π as

$$m_\pi = \lim_{\varepsilon \to 0} \frac{m_\pi^\varepsilon}{\varepsilon}.$$

With this definition, we remove dependence on ε. Therefore, this quantity measures the maximum variation.

The remaining problem is how to evaluate that supremum. For the following quantile class Γ_Q of probability measures and the ε-contamination neighborhood $\Gamma_\varepsilon \subseteq \Gamma_Q$, Theorem 1 provides an algorithm to compute the supremum through the solution of linear programming problems. The quantile class is given by

$$\Gamma_Q = \{r : \underline{p}_j \leqslant r(A_j) \leqslant \bar{p}_j,\ j = 1, \ldots, n\},$$

where $A_1, \ldots, A_n$ is a measurable partition of Θ with $\sum_{j=1}^n \underline{p}_j \leqslant 1 \leqslant \sum_{j=1}^n \bar{p}_j$. The ε-contamination neighborhood $\Gamma_\varepsilon \subseteq \Gamma_Q$ is

$$\Gamma_\varepsilon = \{\pi' = \pi + \eta(q - \pi),\ q \in \Gamma_Q,\ 0 \leqslant \eta \leqslant \varepsilon\}.$$

Theorem 1. *When the probability model does not include data, given $\pi \in \Gamma_Q$ and $\Gamma_\varepsilon \subseteq \Gamma_Q$, then:*

$$m_\pi^\varepsilon = \sup_{\pi' \in \Gamma_\varepsilon} |\dot{U}_\pi(\pi' - \pi)| = \max\{H_1 - U(\pi), U(\pi) - H_2\}\varepsilon,$$

where H_1 and H_2 are, respectively, the optimal values of the linear programming problems:

$$\begin{array}{lll} \max & \sum_{j=1}^n p_j \bar{h}_j, & \min \quad \sum_{j=1}^n p_j \underline{h}_j, \\ \text{s.t.} & \sum_{j=1}^n p_j = 1, & \text{s.t.} \quad \sum_{j=1}^n p_j = 1, \\ & \underline{p}_j \leqslant p_j \leqslant \bar{p}_j,\ j = 1, \ldots, n, & \underline{p}_j \leqslant p_j \leqslant \bar{p}_j,\ j = 1, \ldots, n, \end{array}$$

with $\bar{h}_j = \sup_{\theta \in A_j} u(a, \theta)$ and $\underline{h}_j = \inf_{\theta \in A_j} u(a, \theta)$.

When the probability model contains data, the prior distribution may be updated by Bayes' theorem. In this case, we may compute the derivative by using a result in [18] and the supremum of the derivative can be calculated as in [21]. Concretely, the calculations can be performed by using Theorems 1 and 3 in [11,21].

2.2. Sensitivity with respect to the utility function

We now study changes in $U(a^*; u, \pi) - U(b; u, \pi)$ as we vary the utility function and the alternative b, i.e., the operators in (4)-(6) are analyzed when both the incumbent optimal decision and the utility change. We focus on a class of utility functions which are typical for a wide range of medical applications, as it is shown in the next section. They include a trade-off between a term related to sampling cost and a term referring to the posterior (predictive) probability of some event of interest, that we aim at detecting. Interesting issues for Bayesian robustness concerning other utility functions can be found in [23].

Specifically, assume that the class of utility functions $\mathcal{U}$ is composed by functions of the type:

$$u(a, \theta) = -rn_a - 1_E(a, \theta),$$

where $n_a \in \mathbb{N}$ is the number of samples under action a, E is an objective event, 1_E is the indicator function for E and r is the ratio between sampling cost and the penalty of underachieving the objective. Then, the expected utility is of the form

$$U_r(a) = -r n_a - f(a),$$

where $f(a)$ is the probability of achieving the objective. In Section 3 we illustrate the values with examples.

We study now the impact of the ratio r in the variation of the optimal alternative, through operators (4)–(6). The first result is immediate.

Proposition 1. *The utility functions belonging to $\mathscr{U}$, where $r > 0$, verify:*

(a) *If a^* is the optimal decision for the initial r, then $\forall s > 0$, $U_s(b) \leqslant U_s(a^*)$ $\forall b$ such that $n_b = n_{a^*}$.*
(b) *If a^* is the optimal decision for the initial r, then $\forall s > r$, $U_s(b) < U_s(a^*)$ $\forall b$ such that $n_b > n_{a^*}$.*
(c) *If a^* is the optimal decision for the initial r, then $\forall s < r$, $U_s(b) < U_s(a^*)$ $\forall b$ such that $n_b < n_{a^*}$.*

Therefore, in order to find out how much r may decrease without changing the optimal decision, we have to consider decisions d such that $n_a = n_{a^*} + k$ $(k \geqslant 1)$. Then, the following result is obtained.

Theorem 2. *The parameter r may be decreased as much as Δr without changing the optimal solution, where*

$$\Delta r = r + \min_{k \geqslant 1} \left(\frac{\min_{b \in D_k} f(b) - f(a^*)}{k} \right)$$

and $D_k = \{a : n_a = n_{a^*} + k\}$.

Similarly, we consider decisions a, with $n_a = n_{a^*} - k$, $(k \geqslant 1)$, to know how much r can be increased without changing the optimal decision. Then, a result analogous to Theorem 2 is:

Theorem 3. *The parameter r may be increased as much as Δr without changing the optimal solution, where*

$$\Delta r = \min_{k \geqslant 1} \left(\frac{\min_{b \in D'_k} f(b) - f(a^*)}{k} \right) - r$$

and $D'_k = \{a : n_a = n_{a^*} - k\}$.

Through this section, we have discussed some general principles and proposed new computational strategies, which we will illustrate with two relatively complex medical decision making problems in the next section. Note that we are using operators (5) and (6), looking for changes in the optimal decision. Moreover, Proposition 1 can be applied to study operator (4).

3. Applications to medical decision making

We use two medical decision making examples which are typical for a wide range of applications in this area. The example in Section 3.1 is typical for applications verifying: (i) a semi-Markov model describing transitions between different stages of a disease; (ii) the utility function combines a sampling cost and a payoff related to some event that is easy to evaluate for any assumed values of future data y and parameters θ. The example in Section 3.2 has some key features of another wide class of problems in medical decision making: (i) a parametric hierarchical model to fit data from previous patients is used; (ii) the model is estimated through MCMC posterior simulation; (iii) the expected utility of any alternative contains one term related to sampling cost and another term related to the posterior (predictive) probability of some event of interest.

Other important examples of medical decision making problems with similar features include optimal scheduling of chronic diseases ([24–29]) and optimal design for implantable heart defibrillators ([30–32]).

3.1. Optimal screening schedules for breast cancer

The problem of optimal screening schedules for a chronic disease as, for example, breast cancer is considered in [27]. The decisions to be made include the age α at which to begin regular screening and the frequency δ of screenings, i.e., $a = (\alpha, \delta)$. A four state semi-Markov process to describe the history of a chronic disease is defined in [27,28,33]. The four states are "disease is absent or present but not detectable" (A), "detectable pre-clinical" (B), "clinical" (C) and "death"(D).

For breast cancer, the following specifications are used in [27]. Let p denote the transition probability from A to B ($1 - p$ is the transition probability from A to D), t_2 is the transition time from A to B, t_3 is the transition time from B to C, and t_4 is the transition time from A to D. Also, $t_1 = 1$ or $t_1 = 0$ shows whether a patient's state changes from A to B or from A to D. We denote with $\theta = (t_1, t_2, t_3, t_4)$ the patients history. For the sake of simplicity, we assume no transitions between B and D are possible. If not detected, every pre-clinical case develops into a clinical case in t_3 years. Let $\hat{f}_2(t_2)$ denote the empirical distribution for t_2 based on historical data, and let y be a patient's age, when entering state B. In [27], it is proposed to use $p = 1/8$ and

$$t_2 \sim \hat{f}_2(t_2),$$
$$t_4 \sim \text{Weibull}(a_4, b_4),$$
$$\log(t_3) \sim N(m_{t_2}, s^2),$$
$$s^{-2} \sim \text{Gamma}(a_3, b_3),$$

with $a_4 = 7.233$, $b_4 = 82.651$, $a_3 = 6.33$, $b_3 = 3.36$, $m_0 = 1.4$, $m_1 = 1.6$, $m_2 = -0.038$, and $m_{t_2} = m_0 - \exp(m_1 + m_2 t_2)$. Fig. 1 shows the probability model.

Deciding on an optimal screening schedule requires specifying in the utility function the trade-off r between the screening cost and the probability of early breast cancer detection. Let E denote this event. Let $n_a(\theta)$ be the number of screenings under the state θ and the schedule a. The utility function is given by

$$u(a, \theta) = -r n_a(\theta) + 1_E(a, \theta), \tag{7}$$

where $1_E(a, \theta)$ is 1 if, for θ, we detect breast cancer early, and 0 otherwise.

For a trade-off parameter $r = 0.001$, the maximal expected utility is achieved at $\alpha^* = 42$, $\delta^* = 1.96$, with value 51.8202. Note that the expected utility value has been scaled by a factor of 1000 in order to avoid numbers of low order of magnitude. We consider uncertainty in the probability model for t_2 and t_3. By using the previous information, and for illustration purpose, we embed the prior distribution into a quantile class with two classes for t_3 and six classes for t_2. We then obtain 12 classes, as shown in Table 1, where for example $P(t_2 \in (40, 60], t_3 \in [0, 1.96]) = 0.19654$, for the basic assessment. Note that t_2 and t_3 are not independent. We define the quantile class Γ_Q, and the corresponding family Γ_ε, by matching the quantiles given in Table 1, i.e., lower and upper bound in the general definition of Γ_Q are tied.

We compute $\bar{h}_j$ and $\underline{h}_j$ for each element in the partition. We obtain $H_1 = \sum_{j=1}^{12} p_j \bar{h}_j = 96.2194$ and $H_2 = \sum_{j=1}^{12} p_j \underline{h}_j = 25.1675$. Moreover, $U(\pi) = 51.8202$. By applying Theorem 1, we obtain:

$$m_\pi^\varepsilon = \sup_{\pi' \in \Gamma_\varepsilon} |\dot{U}_\pi(\pi' - \pi)| = \max\{96.2194 - 51.8202, 51.8202 - 25.1675\}\varepsilon = 44.3992\,\varepsilon,$$

and, therefore, $m_\pi = 44.3992$. We might conclude that there is no robustness in expected utility, since the value m_π is large when compared with the value of $U(\pi)$. This lack of robustness confirms the discussion on optimal screening designs for breastcancer. The expected utility surface is extremely flat in a relatively wide neighborhood of the optimal solution, leaving opportunity for extensive disagreement with the optimal schedule. To further understand the nature of this lack of robustness, we consider a reduced quantile class defined by refining the partition.

In this case:

$$m_\pi = H_1 - U(\pi) = \sum_{j=1}^{12} p_j \bar{h}_j - U(\pi) = 44.3992.$$

After observing the values $p_j \bar{h}_j$, it is deduced that the most influential variable is t_3. Then, we split the intervals for t_3, as in [34], becoming now (0,1.80], (1.80,1.96], (1.96,4], and (4,80.18]. New probabilities are assigned for the subintervals, and then, the new value of m_π is 26.3005, which is a considerable reduction. The process can be repeated to obtain more reduction in the sensitivity measure by using the expert opinion information.

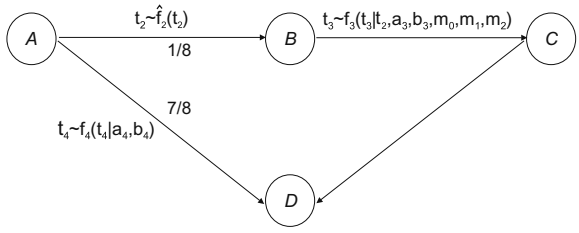

Fig. 1. Probability model for transitions among the four states.

Table 1
Probabilities p_i of the partition for the prior distribution

p_i	$t_3 \in [0, 1.96]$	$t_3 \in (1.96, 80.18]$
$t_2 \in (0, 20]$	0.00643	0.00040
$t_2 \in (20, 30]$	0.02603	0.00241
$t_2 \in (30, 40]$	0.06070	0.01447
$t_2 \in (40, 60]$	0.19654	0.19895
$t_2 \in (60, 80]$	0.13023	0.27854
$t_2 \in (80, 105]$	0.02130	0.06400

3.2. Optimal apheresis schedules

Optimal apheresis designs for cancer patients undergoing high-dose chemotherapy is considered in [35]. Between a pre-treatment and the start of chemotherapy, stem cells are collected to allow for later reconstitution of white blood cell components. There are two possible treatments. Depending on which one the patient undergoes, the first stem cell collection process (apheresis) is scheduled on the fifth or seventh day after pre-treatment. The problem is to decide for which of the remaining days further aphereses should be scheduled. Clearly, the optimal solution should propose stem cell collections on days with high predicted stem cell concentrations. The prediction is based on observations of stem cell levels (represented by CD34 antigen levels) from past patients.

Let y_{ij}, $i = 1,\ldots,I$ and $j = 1,\ldots,n_i$ denote the observed CD34 count for patient i on day t_{ij}. Let $y_i = (y_{i1},\ldots,y_{in_i})$ and $y = (y_1,\ldots,y_I)$ denote the ith patient's data and the combined data vector. A sampling model for the data is specified in [35]. For a new patient, $h = I+1$, let $y_h = (y_{h1},\ldots,y_{hn_h})$ be the (unknown) stem cell counts on days $t_{h1},\ldots,t_{hn_h}$. Denote with $a = (d_1,\ldots,d_{n_h})$ a vector of indicators, with $d_j = 1(0)$ if a stem cell collection is (not) scheduled for day t_{hj}. For the first day t_{h1}, there is a count, y_{h1}, already available. Let $n_a = 1 + \sum_{j=2}^{n_h} d_j$ denote the number of scheduled apheresis, where the first 1 reflects the apheresis at day t_{h1} already realized. Let A be the event of failing to collect a target number y^* of stem cells, i.e., $A = \{\sum_{j=1}^{n_h} d_j y_{hj} L_h < y^*\}$, where L_h is the volume of blood processed at each stem cell collection for the new patient (y_{ij} is recorded per volume unit). The utility function is

$$u(a, y_h) = -c_1 n_a - c_2 1_A(a, y_h), \tag{8}$$

where 1_A is the indicator function for event A, c_1 is the cost of each apheresis, c_2 is the penalty for under-achievement of the target y^* and n_a is the number of apheresis under alternative a. The decision making problem is then:

$$a^* = \arg\max_a \int_\Theta u(a, y_h) p(y_h|\theta)\pi(\theta|y)\mathrm{d}\theta, \tag{9}$$

where $\pi(\theta|y) \propto l(y|\theta)\pi(\theta)$ is the posterior distribution on the unknown model parameters given the observed data y.

In [35], (9) is solved for a particular model $p(y|\theta)$ based on a rescaled Gamma curve for the mean profile of each patient and a hierarchical prior probability model. A covariate, $x_i = 1$ or $x_i = 2$, records each patient's pre-treatment. The two pre-treatments define the choice of the prior distribution at the first level. The hyperprior at the second level is common for both pre-treatments. The model is given by

$$\begin{aligned}
&y_{ij} = z_i g(t_{ij}; e_i, s_i) + \epsilon_{ij}, \quad i = 1,\ldots,I,\ j = 1,\ldots,n_i,\\
&\epsilon_{ij} \sim N(0, \sigma^2),\\
&\theta_i \sim N(\eta_{x_i}, V),\\
&\theta_i = (z_i, e_i, s_i),\\
&\eta_{x_i} \sim N(\mu, \Sigma),\\
&V^{-1} \sim \text{Wishart}(q, (qQ)^{-1}),\\
&\sigma^{-2} \sim \text{Gamma}(a_0/2, b_0/2).
\end{aligned} \tag{10}$$

Here $g(t; e, s) = \Gamma(t; \alpha, \beta)/c$ is the density function of a Gamma distribution with parameters $\beta = e/s^2$ and $\alpha = e \cdot \beta$, chosen to have mean and variance matching e and s^2 and rescaled by $c = [(\alpha-1)/\beta]^{\alpha-1}\exp[-(\alpha-1)]$ such that $\sup(g) = 1$. The rescaling factor is z_i. See Fig. 2 for a schematic representation of model (10).

By using MCMC posterior simulation, the optimal design a^* for a future patient in model (10) with respect to the utility function (8) with $c_1/c_2 = 0.1$ can be found. As an example, for a patient undergoing pre-treatment 1, the optimal apheresis schedule was found to be $a^* = (1,1,0,0,0,0,0)$, with expected utility $U(a^*) = -2.03$.

We consider patients 24 and 25 and apply the results in Section 2.2. Patient 24 begins the stem cell collection on the fifth day after pre-treatment with $L_h = 20$ and $y^* = 200$. For patient 25, stem cell collection begins on the seventh day after the pre-treatment. Again, in this case we have the same values for L_h and y^*. Tables 2 and 3 summarize the best five solutions for patients 24 and 25, respectively.

By using Theorem 2 for patient 24, we obtain $\Delta r = 0.09795$. The current value of r is 0.1, then we can decrease r by 97.95% without changes in the optimal alternative. If we take $r = 0.1 - 0.09795$, we find that alternatives a_4 and a_5 become optimal. Note that in the original utility function we used $c_1 = 1$ and $c_2 = 10$ instead of r, with $r = c_1/c_2$. In this case, c_2 could increase from 10 to 490 without changes in the optimal solution. On the other hand, applying Theorem 3, we find that r could increase by 0.4040816 (404.08%) without changes in the optimal solution. For values of r greater than 0.5040817, the optimal solution indicates stem cell collection only on the fifth day. We conclude that the solution is fairly robust for changes in the utility function for patient 24.

We repeat the study for patient 25, obtaining different results. The parameter r can decrease only by 4.08% without changing the optimal solution. Then, alternative a_2 becomes optimal. However, r could be increased by 165%. In this case, the optimal alternative is to collect stem cells on days seventh and ninth, with expected utility equal to -8.91. This alternative is not among the best five ones. The key conclusion is that the optimal solution is robust against increasing changes and sensitive for decreasing changes.

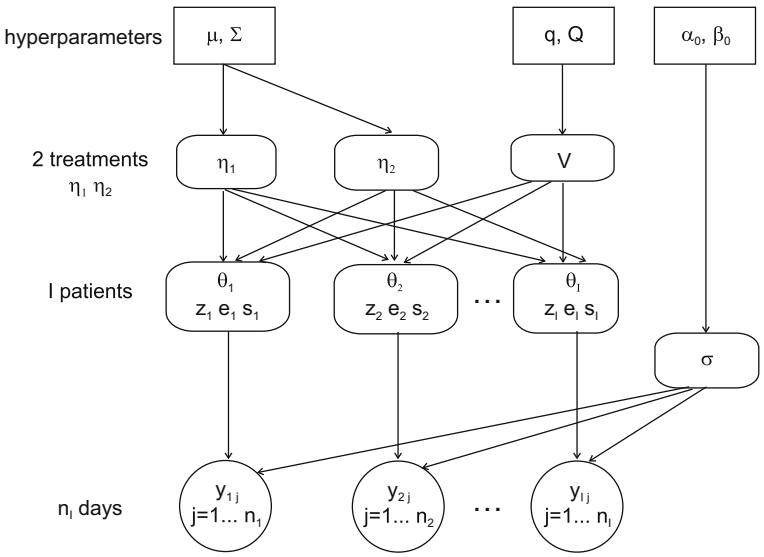

Fig. 2. Graphical model (10). Rectangular boxes correspond to fixed hyperparameters. Rounded boxes are unknown parameters. Circles are observable data. Arrows show conditional dependence.

Table 2
Top five solutions for patient 24

Decision	Days to collect	$U(\cdot)$
a_1	5th and 6th	−2.03
a_2	5th and 7th	−2.08
a_3	5th and 8th	−2.59
a_4	5th, 6th and 7th	−3
a_5	5th, 6th and 8th	−3

Table 3
Top five solutions for patient 25

Decision	Days to collect	$U(\cdot)$
a_1	7th, 9th and 10th	−5.102
a_2	7th, 9th, 10th and 11th	−5.143
a_3	7th, 10th and 11th	−5.224
a_4	7th, 9th, 10th and 12th	−5.429
a_5	7th, 9th, 11th and 12th	−5.429

The practical implication is that the one could comfortably recommend collection on days 5th and 6th, and on days 7th, 9th and 10th, respectively for patients 24 and 25. Also a clinician might choose for additional collection on day 11th, if desired, for patient 25. Although the recommendation is based on a formal decision theoretical setup, with a specific probability model and loss function, the conclusion is more general. One particular feature that traditionally prevents many researchers from applying formal decision theoretical approaches in biomedical problems is the need for a specific utility (or loss function). It is often not clear who is the relevant decision maker and whose loss function should be used, how should monetary costs and health benefits be traded off, etc. The proposed study of robustness and sensitivity mitigates some of these concerns and can facilitate increased use of decision theoretical methods in biomedical applications.

Finally, note that the same MCMC outputs obtained to approximate the optimal design have been used in the sensitivity analysis. This type of analysis has a low computational cost (see, for example, [36]).

4. Conclusion

The Bayesian approach provides a coherent methodology for decision making under uncertainty. In this context, robust analysis should be incorporated to find out how the output of the model changes with variations in the inputs. In this paper, we have discussed some general principles referring to some of the most important works in the field of Bayesian robustness.

New computational strategies have been proposed and their applicability has been illustrated in the context of two relatively complex medical decision making problems. The special importance of decisions in the medical context makes this kind of analysis necessary.

When the complexity of the model increases, the sensitivity analysis becomes also more involved. For example, sensitivity analysis in MCMC methods is a difficult task demanded by several authors. Particularly relevant is the fact that MCMC simulations can be re-used to estimate the sensitivity measures of the proposed approaches avoiding the need for further sampling.

Acknowledgements

The authors would like to thank David Ríos for his constructive comments. This work has been partially supported by *Ministerio de Educación y Ciencia*, Spain (Projects TSI2004-06801-C04-03 and TSI2007-66706-C04-02).

Appendix A. Proofs

Proof 1 (*of Theorem 1*). If $\pi\prime = \pi + \eta(q - \pi)$ then:

$$|\dot{U}_\pi(\pi' - \pi)| = |\dot{U}_\pi(\eta(q - \pi))| = |U(\eta(q - \pi))| = \eta|U(q) - U(\pi)|,$$

therefore,

$$m_\pi^\varepsilon = \sup_{\pi' \in \Gamma_\varepsilon} |\dot{U}_\pi(\pi' - \pi)| = \sup_{q \in \Gamma_Q, \eta \leqslant \varepsilon} \eta|U(q) - U(\pi)| = \varepsilon \sup_{q \in \Gamma_Q} |U(q) - U(\pi)| = \varepsilon \max\left\{ \sup_{q \in \Gamma_Q} (U(q) - U(\pi)), - \inf_{q \in \Gamma_Q} (U(q) - U(\pi)) \right\}.$$

We need to compute $\sup_{q \in \Gamma_Q} U(q)$ and $\inf_{q \in \Gamma_Q} U(q)$. For the supremum, we have, for s>0:

$$\sup_{q \in \Gamma_Q} U(q) = \sup_{q \in \Gamma_Q} \int_\Theta u(a, \theta) q(\theta) \mathrm{d}\theta = \sup_{q \in \Gamma_Q} \sum_{j=1}^{n} \int_{A_j} u(a, \theta) q(\theta) \mathrm{d}\theta,$$

and for all $q \in \Gamma_Q$, we have, $j = 1, 2, \ldots, n$

$$\int_{A_j} u(a, \theta) q(\theta) \mathrm{d}\theta \leqslant q(A_j) \sup_{\theta \in A_j} u(a, \theta).$$

Since $\underline{p}_j \leqslant q(A_j) \leqslant \overline{p}_j$, then $\int_\Theta u(a, \theta) q(\theta) \mathrm{d}\theta \leqslant H_1$. Moreover, $\exists q \in \Gamma_Q$ such that $U(q) = H_1$. Analogously, the infimum is obtained. So, we have $m_\pi^\varepsilon = \max\{H_1 - U(\pi), U(\pi) - H_2\}\,\varepsilon$. □

Proof 2 (*of Proposition 1*). We prove (a) and (b). (c) is similar to (b).

(a) Let b be such that $n_b = n_{a^*}$. Since a^* is optimal for the initial r, we have, for s>0;

$$-r n_b - f(b) \leqslant -r n_{a^*} - f(a^*) \iff -f(b) \leqslant -f(a^*) \iff -s n_b - f(b) \leqslant -s n_{a^*} - f(a^*) \iff U_s(b) \leqslant U_s(a^*)$$

(b) Let b be such that $n_b > n_{a^*}$. Then, n_b can be represented as $n_b = n_{a^*} + k$. If $s > r$ then:

$$\begin{aligned}
&-r n_b - f(b) \leqslant -r n_{a^*} - f(a^*) \iff -r(n_{a^*} + k) - f(b) \leqslant -r n_{a^*} - f(a^*) \iff \\
&-rk - f(b) \leqslant -f(a^*) \iff -sk - f(b) < -f(a^*) \iff -s n_{a^*} - sk - f(b) < -s n_{a^*} - f(a^*) \\
&-s(n_{a^*} + k) - f(b) < -s n_{a^*} - f(a^*) \\
&-s n_b - f(b) < -s n_{a^*} - f(a^*) \iff U_s(b) < U_s(a^*) \qquad \square
\end{aligned}$$

Proof 3 (*of Theorem 2*). By Proposition 1, if r decreases, a^* is preferred to any alternative b with $n_b \leqslant n_{a^*}$. Let b be such that $n_b = n_{a^*} + k$. Alternative b is preferred to a^* with a trade-off $r - \Delta r$ if and only if:

$$-(r - \Delta r)(n_{a^*} + k) - f(b) > -(r - \Delta r) n_{a^*} - f(a^*) \iff \Delta r > \frac{f(b) - f(a^*)}{k} + r$$

As we look for the minimum Δr, we obtain the proposed result. □

Proof 4 (*of Theorem 3*). Analogous to the previous proof. □

References

[1] J.O. Berger, Statistical Decision Theory and Bayesian Analysis, Springer, 1985.
[2] J.M. Bernardo, A. Smith, Bayesian Theory, Wiley, 1994.

[3] S. French, D. Ríos Insua, Statistical Decision Theory, Arnold (2000).
[4] J.O. Berger, An overview of robust Bayesian analysis, Test 3 (1994) 5–124.
[5] D. Ríos Insua, F. Ruggeri, Robust Bayesian Analysis, Springer, 2000.
[6] F. Ruggeri, D. Ríos-Insua, J. Martín, Robust Bayesian analysis, in: D. Dey, C.R. Rao (Eds.), Bayesian Thinking, Modeling and Computation, Handbook of Statistics, vol. 25, North-Holland, 2005, pp. 623–665.
[7] L. Wasserman, Recent methodological advances in robust Bayesian inference, in: J.M. Bernardo, M.H. De Groot, D.V. Lindley, A.F.M. Smith (Eds.), Bayesian Statistics, vol. 4, Oxford University Press, 1992, pp. 483–502.
[8] S. Sivaganesan, Robust Bayesian diagnostics, Journal of Statistical Planning and Inference 35 (1993) 171–188.
[9] D. Dey, S. Ghosh, K. Lou, On local sensitivity measures in Bayesian analysis (with discussion), in: J.O. Berger, B. Betro, E. Moreno, L.R. Pericchi, F. Ruggeri, G. Salinetti, L. Wasserman (Eds.), Bayesian Robustness, IMS Lecture Notes Monograph Series, vol. 29, IMS, 1996, pp. 21–39.
[10] D. Ríos Insua, Sensitivity Analysis in Multiobjective Decision Making, Springer, 1990.
[11] J. Martín, D. Ríos-Insua, F. Ruggeri, Issues in Bayesian loss robustness, Sankhya, Serie A 60 (1998) 405–417.
[12] J. Martín, D. Ríos-Insua, F. Ruggeri, Joint sensitivity in Bayesian decision theory, Test 12 (1) (2003) 173–194.
[13] J. Greenhouse, L. Wasserman, A practical robust method for Bayesian model selection: a case study in the analysis of clinical trial, in: J.O. Berger, B. Betro, E. Moreno, L.R. Pericchi, F. Ruggeri, G. Salinetti, L. Wasserman (Eds.), Bayesian Robustness, IMS Lecture Notes Monograph Series, vol. 29, IMS, 1996, pp. 41–58.
[14] B.P. Carlin, M.E. Pérez, Robust Bayesian analysis in medical and epidemiological settings, in: D. Ríos Insua, F. Ruggeri (Eds.), Robust Bayesian Analysis, Springer, 2000, pp. 351–372.
[15] J. Kadane, C. Srinivasan, Comment on "An overview of robust bayesian analysis" by Berger, Test 3 (1994) 93–95.
[16] C. Goutis, Ranges of posterior measures for some classes of priors with specified moments, International Statistical Review 62 (2) (1994) 245–256.
[17] B. Betrò, A. Guglielmi, Methods for global prior robustness under generalized moment conditions, in: D. Ríos Insua, F. Ruggeri (Eds.), Robust Bayesian Analysis, Springer, 2000, p. 273294.
[18] P. Diaconis, D. Freedman, On the consistency of Bayes estimates, Annals of Statistics 14 (1986) 1–67.
[19] A. Cuevas, P. Sanz, On differentiability properties of Bayes operators, in: J.M. Bernardo, M.H. DeGroot, D.V. Lindley, A.F.M. Smith (Eds.), Bayesian Statistics, vol. 3, Oxford University Press, 1988, pp. 569–577.
[20] F. Ruggeri, L. Wasserman, Infinitesimal sensitivity of posterior distributions, Canadian Journal of Statistics 21 (1993) 195–203.
[21] J. Martín, D. Ríos-Insua, Local Sensitivity Analysis in Bayesian Decision Theory (with discussion), in: J.O. Berger, B. Betro, E. Moreno, L.R. Pericchi, F. Ruggeri, G. Salinetti, L. Wasserman (Eds.), Bayesian Robustness, IMS Lecture Notes Monograph Series, vol. 29, IMS, 1996, pp. 121–137.
[22] P. Gustafson, Local robustness of posterior quantities, in: Ríos Insua, Ruggeri (Eds.), Robust Bayesian Analysis, Springer, 2000, pp. 71–88.
[23] F. Ruggeri, J. Martín, D. Ríos-Insua, Optimal alternative robustness in Bayesian decision theory, Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales, Serie A: Matemáticas 17 (3) (2003) 377–382.
[24] Council on Scientific Affairs. Mammographic screening in asymptomatic women aged 40 years and older. Journal of American Medical Association, 261:2535–2542, 1989.
[25] D.M. Eddy, V. Hasselblad, W. McGivney, W. Hendee, The value of mammography screening in women under age 50 years, Journal of the American Medical Association 259 (10) (1988) 1512–1519.
[26] A.B. Miller, Is routine mammography screening appropriate for women 40–49 years of age?, American Journal of Preventive Medicine 7 (1991) 55–62.
[27] G. Parmigiani, On optimal screening ages, Journal of the American Statistical Association 88 (1993) 622–628.
[28] G. Parmigiani, M.S. Kamlet, Cost-utility analysis of alternative strategies in screening for breast cancer, in: C. Gatsonis, J. Hodges, R.E. Kass, N. Singpurwalla (Eds.), Bayesian Statistics in Science and Technology: Case Studies in Bayesian Statistics, Springer, 1993, pp. 390–402.
[29] G. Parmigiani, Modelling in medical decision making: a Bayesian approach, Statistics in Practice, Wiley, 2002.
[30] R.A. Malkin, T.C. Pilkington, D.S. Burdick, D.K. Swanson, E.E. Johnson, R.E. Ideker, Estimating the 95% effective defibrillation dose, IEEE Transactions on Biomedical Engineering 40 (3) (1993) 256–265.
[31] A.S. Manolis, W. Tan-DeGuzman, M.A. Lee, H. Rastegar, C.I. Haffajee, S.K. Haung, Clinical experience in seventy-seven patients with automatic implantable cardioverter defibrillator, American Heart Journal 118 (1989) 445–450.
[32] M. Clyde, P. Müller, G. Parmigiani, Optimal for heart defibrillators, in: D.A. Berry, D.K. Stangl (Eds.), Bayesian Biostatistics, Marcel Dekker, 1994, pp. 297–320.
[33] M. Zelen, Optimal scheduling of examinations for the early detection of disease, Biometrika 80 (2) (1993) 279–293.
[34] S. Ríos Insua, J. Martín, F. Ríos Insua, D. Ruggeri, Bayesian forecasting of accident proneness evaluation, Scandinavian Actuarial Journal 2 (1999) 134–156.
[35] J. Palmer, P. Müller, Bayesian optimal design in population models of hematological data, Statistics in Medicine 17 (1998) 1613–1622.
[36] C.J. Pérez, J. Martín, M.J. Rufo, MCMC-based local parametric sensitivity estimations, Computational Statistics and Data Analysis 51 (2006) 823–835.